Biological Programmable Logic Device in *Escherichia coli*

Mingyang Sun

A Thesis

In

The Department

Of Biology

Presented in Partial Fulfillment of the Requirements

for the Degree of Master of Science at Concordia University

Montréal, Québec, Canada

April, 2012

CONCORDIA UNIVERSITY

School of Graduate Studies

This is to certify that the thesis prepared

By: MINGYANG SUN

Entitled: Biological Programmable Logic Device in Escherichia coli

and submitted in partial fulfillment of the requirements for the degree of

Master of Science in Biology

complies with the regulations of the University and meets the accepted standards with respect to originality and quality.

Signed by the final examining committee:

Dr. Vladimir Titorenko Chair

Dr. Patrick Gulick Examiner

Dr. Vincent Martin Examiner

Dr. Luc Varin Supervisor

Approved by Dr. Selvadurai Dayanandan

Chair of Department or Graduate Program Director

Brian Lewis

Dean of Faculty

Date April 13th 2012

Biological Programmable Logic Device in *Escherichia coli*

Mingyang Sun

Abstract:

Bio-computing and Boolean logic gates simulation and implementation in bacteria and eukaryotic cells is currently a hot subject in Synthetic Biology. Computer chips relies on combinatorial Boolean logic circuits, which are the combination of basic Boolean logic gate operations to function (AND, OR, NOT) and these circuits are the basic components for computer systems. Hence constructing artificial gene modules implementing basic Boolean logic operations is the first crucial step towards building a complex bio-computer. Currently, basic logic gene networks have been constructed using different genetic components and tested. However, these studies didn't provide a full set of synthetic biological parts, which could be assembled later in one system to fully exploit the computing ability of combinatorial circuits. In this project, a set of logic gene modules (A AND B, A' AND B, A AND B', A' AND B') responding to two different variable inputs (IPTG and anhydro-tetracycline) and their variants were constructed and tested. These modules could potentially be integrated into a single system to provide all possible 16 different outputs from two variable inputs. In this thesis, the four modules were designed; constructed and tested for their logic functions. The positive and negative results presented in this thesis provide some insights on the design principles for creating artificial logic gene parts to fulfill the needs for building higher-level bio-computing machines.

Acknowledgement

Foremost, I would like to express my sincere gratitude to my advisor Dr. Luc Varin for the continuous support of my master study and research, for his patience, enthusiasm, and immense knowledge. His guidance helped me in all the time of research and writing of this thesis. I could not have imagined having done my master degree without the help from Luc. Besides my advisor, I'd like to thank my co-supervisor/committee: Dr. Nawwaf Kharma for being a great mentor and friend who has been constantly giving me advises and encouragement. Also my second committee Dr. Martin, who has been supporting me by guiding me to his knowledgeable researchers such as Andy Ekins, Corinne Cluis and others for consultant and help.

Of course, lots of love and support was coming from my parents and family. They have been my strongest support since I started this project and saw me through with no doubt. Therefore I'd love to express my dearest thank to them!

Last but not least, I'd like to shout out loud to all my dear friends from the lab:
Samar Elzein, Juan Francisco Leyva, Saddiqa, Samuel Cournoyer, Yao Zhang, Mosha and those from the entire Biology department…I have had the most amazing 3 years of my life in this department and in Montreal…It is all because of you guys. You guys have NO IDEA how much you guys mean to me…to sum it all…that was hell of a ride and you guys are my best friends for life.

TABLE OF CONTENTS

List of Abbreviations

aTc: anhydro-tetracycline

BIO-PLD: Biological Programmable Logic Device

BLG: Boolean Logic Gates

GRN: Gene Regulatory Network

GFP: Green Fluorescent Protein

IPTG: Isopropyl β-D-1-thiogalactopyranoside

OD: Optical Density

PCR: Polymerase Chain Reaction

RBS: Ribosomal Binding Site

RNAP: RNA Polymerase

SDM: Site Directed Mutagenesis

SD sequence: Shine-Dalgarno sequence

Tet: Tetracycline

TF's: Transcription Factors

List of Figures

List of Tables

Chapter I – Introduction

I-1. Prologue

The number of living organisms in nature is an astronomical figure. Every species and every individual represents a unique morphological and genetic variant in its own group. The study of genetic similarities allows linking different organisms and classifying them into groups in a sophisticated system. More importantly, the great genetic diversity and variations are not only fingerprints, which allow us to distinguish one individual from another, but also a great gift of nature that offers the possibility to use or modify their genes to create novel functions. The accumulation of knowledge on gene function and regulation opens up the horizon for the forward engineering of novel and complex artificial gene networks. This is the essence of Synthetic Biology.

When we look back four decades ago, a visionary talk given by Richard Feynman had addressed for the first time the desire to build a computing machine with *sub-microscopic* sized biological components [1]. This talk from a physicist ignited a new wave of research in Molecular Biology. The concept of modern Synthetic Biology started to emerge when Dr. Waclaw Szybalski described what he thought Synthetic Biology should be in his book in 1974 [2]. Over the past two decades, together with the growing understanding of natural gene networks in prokaryotic and eukaryotic organisms, Synthetic Biology has evolved with an unparalleled speed. The way Synthetic Biology develops reminisces the early age of Electrical Engineering, which branched out from physics to form a unique discipline by itself. Recently, more and

more Synthetic Biology designs have embraced Engineering concepts, e.g. interoperable, well-specified modules and computer-aided design [3], which help greatly to build the foundation of this exciting new field.

I-2 History and current research in Synthetic Gene Regulatory Network

Over the past decade, biologists and engineers have created a variety of synthetic gene devices with different functionalities, from the first bacterial toggle switch, which was constructed from two repressible promoters arranged in a mutually inhibitory network [4] to genetic switches [5-11], digital logic evaluators [12,13], and *in vitro* or *in vivo* biocomputing devices with 1 or 2 input combinatorial Boolean Logic gates [14-17] Some of these devices have shown valuable potential in gene therapy [3], as biosensors [18-20], cancer-targeting bacteria [21] and drug delivery systems[22].

The ultimate engineering goal in Synthetic Biology is to create a fully controllable artificial Gene Regulatory Network (GRN). Artificial GRN feature a programmable gene network with predictable dynamic behavior. The purpose of creating artificial GRN is to allow dynamic and controllable orchestration of gene expression for the users to provide inputs that will be translated into predictable outputs. In order to create such a programmable gene network, we require well-defined modules with interoperable design with highly reliable and stable relationship between input and output, and a robust response from the cells. However, in most application-driven cases, engineered organisms contain only simple gene circuits that do not fully exploit the potential of synthetic biology. A fundamental gap between basic genetic

circuits and the promise of assembling these circuits into more complex gene networks that exhibit robust, predictable behaviours is the biggest hurdle we have to conquer.

I-2-1. Combinatorial Logic Gene circuitry and Bio-computing machines

Among the broad spectrum of synthetic GRN a major category of circuitries is based on combinatorial logic. Such a circuit features multiple input signals (chemical or physical signals) controlling gene expression. In all cases, the combined effect of input signals usually results in a regulatory logic that resembles simple or complex electrical logic circuits, hence possessing the great potential to solve complex logic problems. A complex GRN can be broken down to simple Boolean Logic gates such as AND, OR and NOT gates. These basic functions are able to read one or two input variables then provide a specific output signal. This highly specific input/output relationship is critical for the proper execution of the logic function. For example, the Boolean logic AND gate only provides a true output (1 or on) if both input variables are true (1). The AND gate never provides a true value if one or two variables are false (Figure 1).

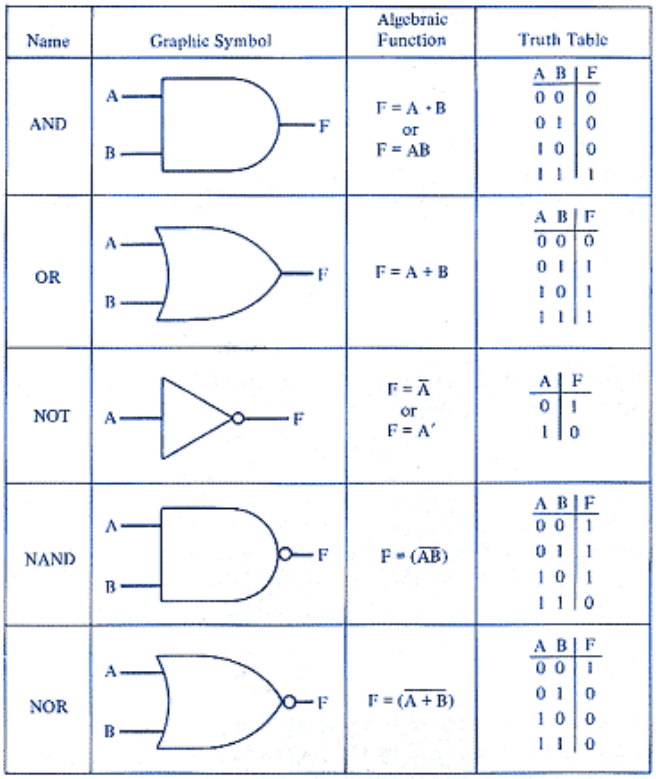

Figure 1. Boolean Logic truth table, A and B represent inputs, F is the output

This type of input/output specific pattern exhibited by Boolean logic gates has been shown to occur in nature [26-28]. In biological systems, gene regulation is highly complex and organized. From Prokaryotes to Eukaryotes, the complexity increases significantly. One or a few signalling molecules sometimes regulate one gene directly or indirectly. In the classic example of the Lac operon in *E. coli*, a chemical "inducer"-lactose and a repressor Lac I, control the transcription of the operon as on or off,. Without the presence of lactose, the operon is shut down as a result of the Lac I repressor binding to the Lac operators (cis-regulatory elements). The repressor/operator interaction prevents the transcription from initiating [29]. When lactose is present, it works as an inducer to set free the Lac I repressor from the operators and makes the transcription of the operon possible [29]. Such regulatory

control mechanism is used to regulate many different genes or operons in prokaryotes.. For example, the arabinose operon and the TetR-controled gene regulation system are using similar control mechanisms [30]. By combining two simple control mechanisms in a novel synthetic promoter, we can create a Boolean logic gene regulatory system implementing the AND gate.

Exploiting the potential of existing natural genes or gene modules is the first step towards creating synthetic logic genes. A lot of efforts have been put into understanding gene regulatory networks that exhibit Boolean logic function patterns, such studies include the cis-regulatory input function at the gene level in *E. coli* [34] or at the genome level [35], and the effect on Transcription Factors (TF's) binding dependences on regulatory functions [37]. A lot of the attention has been drawn to the potential and limitation of combinatorial signal integration [38] and building combinatorial library that contains a variety of promoters responding to different inputs [36] as well.

Some logic network designs include the construction of a simple AND gate responding to IPTG and arabinose in *E. coli* [16], a biosensor that can integrates multiple environmental signal inputs [14], a gene with interchangeable components, that can achieve 11 out of the total 16 different logic outputs with two input signals [25] and a modular orthogonal AND gate with a novel hetero-regulation module [24]. Similar work has been done in Eukaryotes as well, such as the RNAi-based logic evaluator that operates in mammalian cells [17], and a set of biologic gates responding to 3 different molecular inputs [33], to mention a few. Some of this work

has valuable real-life applications, like the Bio-computer that can identify and analyse mRNA of disease-related genes [32] and the cancer cell-invading bacteria [21]. All these results show that Boolean logic gates can be implemented using GRN system. However most of the combinatorial Boolean logic networks built to date are simple logic gates or combination of 2 or few gates that only allows computing simple functions. Furthermore, once implemented in a cell, the GRN cannot be easily re-configured to perform other tasks.

I-2-2. Working on the foundation

In order to build a functional logic synthetic GRN, it is important to decipher how regulatory elements i.e. TFs interact with each other, i.e. binding independently, competitively, orderly or jointly [37]. It is also crucial to understand the relative strength of a wide range of regulatory elements, i.e. operators, -35, -10 boxes, SD sequence etc. Furthermore, how cell physiology is affected by these gene regulations plays an important role in designing functional logic synthetic GRN.

We need more foundation work to provide construction guidelines and a library of biological parts for creating genes that respond to more than one artificial regulatory signal. Furthermore, we need to create modular synthetic gene networks that can be used in different cellular context. In order to successfully create a precisely regulated synthetic gene, researchers utilize combinatorial synthesis [30] and rational design [39] as two main approaches and through which they find novel functionality they desire. Meanwhile, more and more synthetic genes are

emphasizing the modular property so they could perform robustly in different cell contexts [24,39]. MIT has established the BioBricks Foundation to solely promote the open development of Synthetic Biology. The BioBricks Foundation has invested a tremendous amount of efforts on building a registry of standard biological parts that can be shared among researchers and to develop novel methodologies that facilitates the construction of novel parts [40].

The potential of Synthetic Biology is still at its very rudimentary stage. The attempts to design and create synthetic gene networks have strongly helped to push this research area forward but also, exposed our need to develop more reliable and predictable biological parts that can be assembled together. Furthermore, most of the up-to-date synthetic GRN are limited to prokaryotes [33].

I-3. Biological Programmable Logic Device Design and Potential Application

The use of chimeric promoters to control gene expression has shown some potential to simulate the behaviour of Boolean Logic Gates [16] and is the basis of my research. What has not been done is to design a series of modules with different logic functions using chimeric promoters to later be integrated into a single GRN and create a more complex logic system that could utilize two input signals to carry out more complex calculations. For example, when a AND gate and a NOT gate are used in the same system, a new NAND gate is created and the possible number of outputs from two input signals is 16. Current research has successfully achieved most of these 16 outputs separately from individual synthetic gene modules [26]. However, it

is still a challenge to have one single artificial GRN providing all 16 outputs from two variables. Combining simple Boolean logic gene modules to achieve a complete set of logic output is significant since the amount of output could increase exponentially with the increase of input signals. Therefore, the potential for Bio computing of such a complex logical gene regulation system is significant.

In the BIO-PLD project presented in this thesis, I attempted to design a programmable artificial Gene Regulatory Network (GRN) with 4 biological logic modules that can solve complex Boolean problems when assembled into one system. Every artificial GRN module in this project is made of either one or two parts/categories; we named the 2 members of the first category as "inverters", each one of them behaving like a "NOT" gate in Boolean logic, and the second category contains 4 individual "min-term" genes. Overall, there were six novel artificial genes designed and produced during this project (Figure 2).

Within the first category, the two inverter genes have either a native Lac promoter or a Tet promoter. These two promoters are repressed by the LacI or TetR repressor and subsequently can be induced by IPTG or anhydro-tetracycline. The promoters are linked to the coding sequence of the yeast Gal4 DNA binding protein or the λ C1 repressor protein was attached to the two promoters. The four min-term genes are constructed with synthetic promoters that respond either directly or indirectly (via inverters) to two different input signals (IPTG and aTc) and produce a common output, the Green Fluorescent Protein (GFP) (figure 2). The GFP is produced only when the correct logic combination of inputs is provided. The 4 "min-

term" genes and the 2 "inverters" make up the 4 basic Boolean logic modules; A&B, (NonA)&B, A&(NonB) and (NonA)&(NonB).

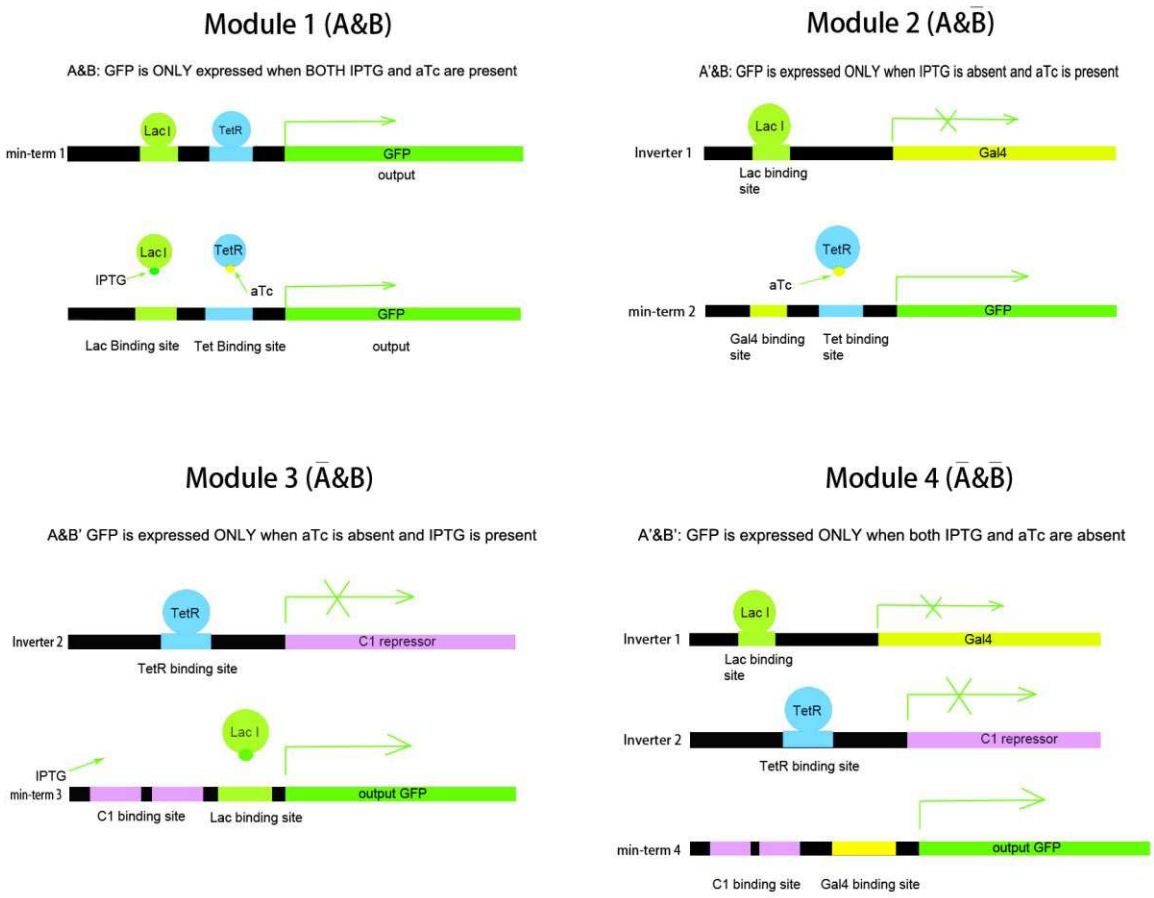

Figure 2. Graphic representation of the four synthetic BIO-PLD modules

The construction and characterization of these modules are the main goals and the first crucial step towards building a sophisticated, fully functional BIO-PLD. The Boolean logic modules constructed in my project are different from the AND or NAND logic gene circuits that have been previously been reported in the literature by having only minimal regulatory sequences (operators) in their regulatory region (promoter). In my work, I constructed different versions of these four modules and

evaluated their response to the two chemical input signals IPTG and anhydro-tetracycline. The results have provided insight on understanding some of the designing rules for creating a logic regulatory circuit and paved the road for future attempts to design such modules and perhaps integrating these modules into a single logic GRN.

Chapter II-Material and Methods

II-1. Materials

All the biological parts used in this study were from the Biobricks Foundation, MIT, (USA).

The *pfu* turbo DNA polymerase and *pfu* turbo buffer were purchased from Agilent-Stratagene. (USA)

All primers were purchased from IDT-DNA (USA) or Biocorp (Montreal)

Mini-prep kits were purchased from Fermentas Inc or BioBasic (CANADA).

All restriction enzymes and T4 DNA ligase were from New England Biolab (USA)

All molecular biology grade chemicals were purchased from Sigma-Aldrich (USA)

E. coli strain DH5αZ1 was kindly provided by Dr. Elaine Newman (Concordia University)

The genotype of DH5αZ1 is $laci^{q}$, PN25-tetR, Sp^{R}, deoR, supE44, Delta (lacZYA-argFV169), Phi80 lacZDeltaM15, hsdR17 (rK- mK+), recA1, endA1, gyrA96, thi-1, relA1

II-2. Methods

A total of six novel synthetic genes were constructed for the BIO-PLD project and these genes were used to create four modules representing four logic gate combinations: A&B, A'&B, A&B' and A'&B'. The six genes were divided into two categories; min-terms and inverters.

The six genes were built from existing biological parts obtained from the BioBricks Foundation (http://biobricks.org/). The four min-terms were constructed using site-directed mutagenesis (SDM) on plasmids containing the proper biological parts. The SDM experiments were done using the Stratagene QuickChange® Site Directed Mutagenesis (SDM) kit and the *pfu* turbo DNA polymerase and associated buffer. A total of nine pairs of mutagenesis primers were designed for creating the four min-terms and their variants.

II-2-1. Construction of min-term 1 (A&B)

The first min-term gene (A&B) is a "AND" logic gate and was constructed by mutating the promoter of the existing BioBricks biological part BBa_I13522 (BioBricks Registry of Standard Biological Parts). The BBa_I13522 biological part contains a Tet promoter that houses the two operators; TetR1 and TetR2, a Shine-Dalgarno (SD) sequence and the GFP coding sequence (Figure 3). The promoter was engineered to a Lac/Tet chimeric promoter that will respond to IPTG and anhydro-tetracycline (aTc) by replacing one of the TetR operator sites with a Lac O1 operator sequence. Four different variants of the Lac/Tet chimeric promoter were created in this project. The four synthetic promoters differ in the localization or in the sequence of the lac operator.

Figure 3: Part BBa_I13522 from BioBricks™ has two Tet operators, TetR1 and TetR2 shown in red in its promoter. (The three underlined parts represent the -35, -10 boxes and part of the Shine-Dalgarno sequence).

Construction of variant 1 of min-term 1

For the first variant, the Tet operator upstream of the -35 box in BBa_I13522 was mutated to the O1 Lac operator having the sequence 5'-aattgtgagcggataacaa-3'. (Figure 4)

Figure 4. Variant 1 of min-term 1 with the Lac O1 sequence (red) replacing the original TetR 1 operator sequence

Primers used for the construction of min-term 1 variant 1:

Forward: 5'- gccgcttctagagaattgtgagcgctcacaattttgacatccctatcag -3' Tm: 67.4 ^{o}C

Reverse: 5'- ctgatagggatgtcaaaattgtgagcgctcacaattctctagaagcggc -3' Tm: 67.4 ^{o}C

After mutating the BBa_I13522 plasmid with the two primers, the mutagenesis product was digested with the restriction enzyme Dpn I at 37 ^{0}C for 1 hr-1.5 hrs, then transformed into DH5αZ1 *E. coli* strain. This strain produces the Lac I repressor (lacI^{q}) and the tetR repressor (PN24-tetR) from the genome, which are required for controlling the synthetic logic gene module.

The transformed DH5αZ1 cells were plated on LB-Spec100/Amp100 plates, and grown at 37°C overnight. About 30 colonies were chosen from the plate for colony PCR to confirm the promoter mutation. The colony PCR primers used to test the first variant of min-term 1 are listed below:

Forward Primer: 5'-aattgtgagcggataac-3' Tm: 49 °C

Reverse Primer: 5'-gtatacatcatggcagaca-3' Tm: 49 °C

Plasmid DNA from four positive colonies was prepared for sequencing. The colonies were grown in 3 ml liquid LB-Spec100/Amp100 medium at 37°C overnight and the plasmids extracted with the Fermentas Miniprep kit. Plasmids were quantified on 1% Agarose gels, and then prepared along with the sequencing primer set (same for all four variants) according to the concentration requirement of the Genome Quebec Sequencing Centre. The plasmids, which were successfully mutagenized, were stored at -20°C and glycerol stocks of the colonies that contain the mutagenized plasmids were stored at -80°C.

Construction of variant 2 of min-term 1

The same procedure was carried out for creating the three other variants of the pLacO/TetRO chimeric promoter. Variant 2 features a Lac O1 operator sequence located between the -35 and -10 boxes of the Tet promoter (Figure 5).

Figure 5. Variant 2 of min-term 1 with the Lac O1 sequence (red) replacing the original TetR 2 operator sequence

Primers used for the construction of variant 2:

Forward Primer:

5'-atcagtgatagagattgacattgtgagcggataacaagatactgagcactactaga-3' TM: 65.8 ^{o}C

Reverse Primer:

5'-tctagtagtgctcagtatcttgttatccgctcacaatgtcaatctctatcactgat-3' TM: 65.8 ^{o}C

Colony PCR primers for variant 2:

Forward Primer: 5'- aattgtgagcggataac -3' Tm: 49 ^{o}C

Reverse Primer: 5'- gtatacatcatggcagaca -3' Tm: 47.3 ^{o}C

Construction of variant 3 of min-term 1

For variant 3, a different sequence of the lac operator (5'-ttgtgagcggataacaa-3') was introduced between the -35 and -10 boxes (Figure 6). It was shown that LacI binds this sequence with a higher affinity as compared to the native Lac O1. [41]

Figure 6. Variant 3 of min-term 1 with the Lac O4 sequence replacing the original TetR 2 operator sequence

Primers used for the construction of variant 3:

Forward Primer:

5’-atcagtgatagagattgacttgtgagcggataacaatgatactgagcactactaga-3’ Tm: 52.7 ^{0}C

Reverse Primer:

5’-tctagtagtgctcagtatcattgttatccgctcacaagtcaatctctatcactgat-3’ Tm: 52.7 ^{0}C

Colony-PCR primers for variant 3:

Forward Primer: 5’-ttgtgagcggataaca-3’ Tm: 49.3 ^{0}C

Reverse Primer: 5’-gtatacatcatggcagaca-3’ Tm: 47.3 ^{0}C

Construction of variant 4 of min-term 1

Variant 4 has another Lac operator sequence (5’-tgtgagcgctcacaatt-3’) residing between the -35 and -10 boxes (Figure 7). Again it was shown that LacI binds to this operator sequence with a higher affinity as compared to the wild type Lac O1 operator [41].

Figure 7. Variant 4 of min-term 1 with the Lac O4S sequence (red) replacing the original TetR 2 operator sequence

Primers used for the construction of variant 4:

Forward Primer:

5’-tcagtgatagagattgacttgtgagcgctcacaattgatactgagcactactaga-3’ Tm: 54 ^{0}C

Reverse Primer:

5’-tctagtagtgctcagtatcaattgtgagcgctcacaagtcaatctctatcactga-3’ Tm: 54 ^{0}C

Colony-PCR primer sequences for variant 4:

Forward Primer: 5’-gtgagcgctcacaat-3’ Tm: 49.5 ^{o}C

Reverse Primer: 5’-gtatacatcatggcagaca-3’ Tm: 47.3 ^{o}C

II-2-2. Construction of min-term 2, 3 and 4

Min-term 2 was constructed using pTet-GFP (BBa_I13522) in the pSB1AK2 plasmid as starting material (same backbone as min-term 1). In min-term 2, one of the Tet operators was replaced by the Gal4 binding sequence (Figure 8).

Figure 8. Min-term 2 construct with the Gal4 binding site replacing the Tet R1 operator

Primers used for the construction of min-term 2:

Forward primer:

5’-ccgcttctagagtcggaggacagtactccgattgacatccctatc-3’ Tm: 69.8 ^{0}C

Reverse primer:

5’-gatagggatgtcaatcggagtactgtcctccgactctagaagcgg-3’ Tm: 69.8 ^{0}C

Min-term 3 was constructed using λ pR-GFP (BBa_S03335) in the pSB1A3 plasmid as starting material. This term was constructed by the insertion of the Lac O1 operator into the λ pR-GFP backbone while keeping both λ pR O1 and O2 operators. The Lac operator was introduced downstream of -10 and upstream of the SD sequence. The final construct of min-term 3 is illustrated in Figure 9.

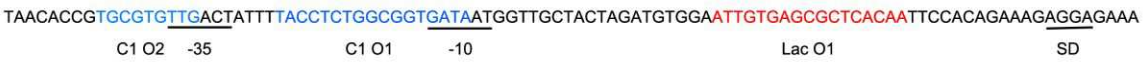

Figure 9. Min-term 3, the red sequence is the Lac O1 operator sequence inserted into the λ pR -GFP promoter.

Primers used for the construction of min-term 3:

Forward Primer:

5'-ggtgataatggttgctactagatgtggaattgtgagcgctcacaattccacagaaagaggagaaatactagatg-3'

Tm: 79 ^{0}C

Reverse Primer:

5'-catctagtatttctcctctttctgtggaattgtgagcgctcacaattccacatctagtagcaaccattatcacc-3'

Tm: 79 ^{0}C

Min-term 4 was created using the same template used for min-term 3. A Gal4 DNA binding sequence was inserted downstream of -10 and upstream of the SD sequence (figure 10).

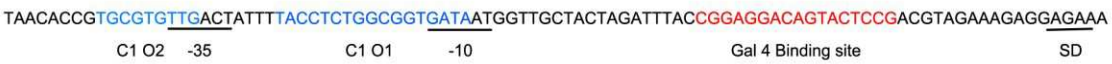

Figure 10. Min-term 4 with the Gal 4 DNA binding site inserted between -10 and the SD sequence.

Primers used for the construction of min-term 4:

Forward Primer:

5'-ggtgataatggttgctactagatttaccggaggacagtactccgacgtagaaagaggagaaatactagatg-3'

Tm: 79.7 ^{0}C

Reverse Primer:

5'- catctagtatttctcctctttctacgtcggagtactgtcctccggtaaatctagtagcaaccattatcacc-3'

Tm: 79.7 ^{0}C

II-2-3. Construction of the two inverters

Modules 2, 3 and 4 have at least one negation of one of the variables in their logic function. The negation of one variable is implemented by the introduction of an inverter repressor-coding gene (Figure 2, Chapter I).

Inverter 1 was made by the fusion of the *E. coli* Lac promoter to the Gal4 DNA binding domain from *Saccharomyces cerevisiae*. The promoter and Gal4 sequences were first amplified separately from two individual BioBricks plasmids (BBa_J04430

and BBa_K105007, respectively) with overlapping primers to connect the two parts together.

The Lac promoter part was amplified from BBa_J04430 with the following primers.

Primer 1: 5'- gaattcgcggccgcttctagagc -3' Tm: 62.4 ^{0}C

Primer2: 5'- gatagaagacagtagcttcatctagtatttctcctcttt -3' Tm: 59.4 ^{0}C

The partial Gal4 sequence was amplified from BBa_K105007 with the following primers.

Primer 3: 5'- aaagaggagaaatactagatgaagctactgtcttctatc-3' Tm: 59.4 ^{0}C

Primer 4: 5'- ctgcagcggccgctactagtacg -3'. Tm: 63.6 ^{0}C

The second and third primers were designed to overlap with each other so they can serve to assemble the two pieces together during PCR.

The PCR was carried out first with two individual reactions aiming at amplifying the two pieces of the inverter for 5 cycles to produce enough fragments with overlapping sequences. Then the amplification was stopped and the two reactions were mixed together to allow the amplification of the full-length synthetic gene. The pLac-Gal4 PCR product was purified by gel-electrophoresis followed by extraction of the DNA band using IBI gel extraction kit. The gel-purified PCR product was ligated into the plasmid p-Jet (From Fermentas) and transformed into *E. coli* DH5αZ1 for sequencing.

The same procedure was applied to create the second inverter gene (pTet-CI). However, in this case two variants differing in the strength of the ribosomal binding site were created. For the first variant, the original ribosome binding site of part BBa_I13522 (from BioBricks part registry) was kept while in the second variant we

replaced it with one having 11.9% of the original strength (BioBricks Foundation, Part:BBa_J61101).

For the construction of the first variant, the Tet promoter sequence was amplified from BBa_I13522 with the following primers.

Primer 1: 5'-gaattcgcggccgcttctagagtccctatc-3' Tm: 64.6 ^{0}C

Primer2: 5'-tgttaatggtttctttttttgtgctcatctagtatttct-3' Tm: 58.7 ^{0}C

The λ CI sequence was amplified from BBa_K105004 with the following primers.

Primer3: 5'-agaaatactagatgagcacaaaaaagaaaccattaaca -3' Tm: 58.7 ^{0}C

Primer 4: 5'- ctgcagcggccgctactagtattattaag -3' Tm: 60.2 ^{0}C

The second and third primers were designed to overlap with each other so they can serve to assemble the two pieces together during PCR.

The resulting plasmid was used to construct the second variant by changing the RBS using site-directed mutagenesis and the following two primers:

Primer 1:

5'-gatagagatactgagcactactagagaaagacatgagttactagatgagcacaaaaaagaaaccattaa-3'

Tm: 78.1 ^{0}C

Primer2:

5'-ttaatggtttctttttttgtgctcatctagtaactcatgtctttctctagtagtgctcagtatctctatc-3' Tm: 78.1 ^{0}C

II-2-4. Construction of complete module 2, 3 and 4

To create module 2, 3 and 4, we ligated the inverters to their corresponding min-terms. Module 2 was created by ligating min-term 2 with the pLac-Gal4 inverter. All

min-terms and the inverter sequences were flanked by EcoR I, XbaI at the 5' end and SpeI, PstI at the 3' end. Following the BioBricks standard assembly protocol (Figure 11), the plasmid containing the inverter pLac-Gal4 was digested with EcoR I and XbaI, hence creating a linear plasmid piece with two incompatible sticky ends. Then min-term 2 was excised out from its plasmid pSB1AK3 with EcoR I and SpeI and gel purified. The two linear pieces are ligated together using T4 DNA ligase. The EcoRI sites anneal together and the XbaI site anneals with the SpeI site. The final plasmid contains the same four restriction sites just like the min-term and the inverter containing plasmid. This type of ligation guarantees all synthetic genes to have the BioBricks standard restriction sites flanking their 5' and 3' extremities.

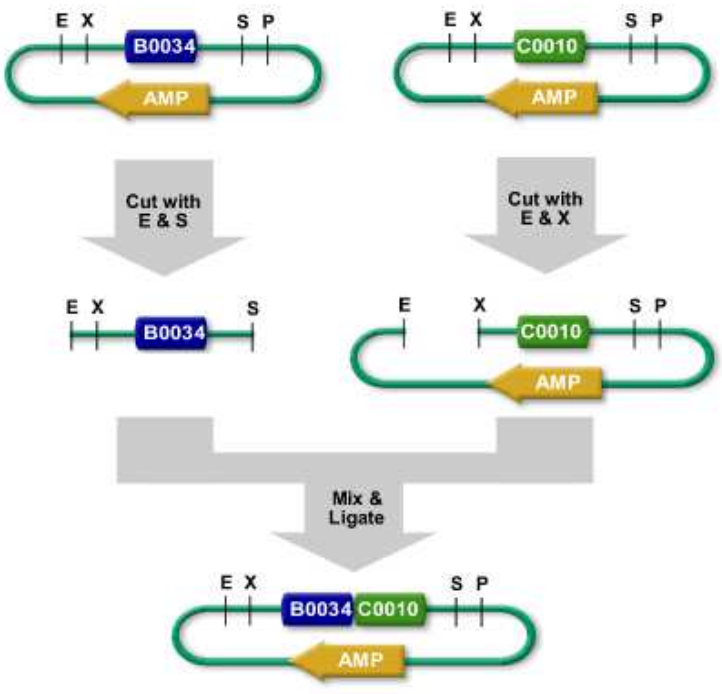

Figure 11. BioBricks Standard Assembly protocol (Figure taken from the BioBricks Foundation web site)

Module 3 was created using the same experimental protocol as described for module 2 but with its own min-term (pCI/Lac-GFP) and its own inverter gene (pTet-CI).

Module 4 contains both inverters (pLac-Gal4 and pTet-CI) and was assembled in a sequential manner downstream of min-term 4. The assembly procedure was the same as described above.

Ligated plasmids containing the entire module 2, 3 and 4 were transformed into *E. coli* DH5αZ1 and plated on to LB-AMP/SPEC 100 Agar plates. The colonies were tested by Colony PCR to confirm that the expected ligation products were present in the transformants.

The sets of primers used for Colony PCR for module 2, 3 and 4 are listed below:

Module 2 Colony PCR primers:

Forward: 5'- cggaggactgtcctcc -3' Tm: 54.3 ^{o}C
Reverse: 5'-ttgtctgccatgatgtatacc-3' Tm: 52.2 ^{o}C

Module 3 Colony PCR primers:

Forward: 5'- aattgtgagcggctcacaatt -3' Tm: 55.6 ^{o}C
Reverse: 5'- gtgtgagttatagttgtattcca -3' Tm: 50.1^{o}C

Module 4 Colony PCR primers:

Forward: 5'- cggaggacagtactccg -3' Tm: 54.2 ^{o}C

Reverse: 5'- gtgtgagttatagttgtattcca -3' Tm: 50.1°C

II-2-5. Induction response tests of module 1 to 4 and statistic tests

In vivo induction tests were done for all modules and their variants to test their logic function. Single colonies containing the different modules were picked and grown overnight in 3ml LB-liquid medium containing 100 ug spectinomycin/ml and 100 ug ampicilin/ml then sub-cultured by adding 100 ul from the 3 ml overnight culture to a flask containing 20 ml of fresh medium. Cultures were grown while shaking 200-250 rpm at 37 °C and the OD was monitored every 20 min. When the OD600 was reaching 0.6, part of the 20 ml culture was divided into four 3 ml aliquots. The first aliquot was used as a control and was not induced with any chemicals. 0.03 ng/ml anhydro-tetracycline dissolved in ethanol was added to the second tube. 1nM of IPTG dissolved in water was added to the third tube. Finally, 0.03 ng/ml anhydro-tetracycline and 1nM IPTG solutions were added to the fourth aliquot. GFP production was monitored 1 hr and 2 hr (in some cases 30 min and 4 hr as well) after induction. 200 ul of the cell culture from each of the four tubes was used for GFP detection and OD600 reading. GFP and OD600 data were collected using a Wallac 1420 multi-label counter. Three to five GFP fluorescence assay were analyzed for the biological replicates.

Statistical average are calculated by using biological replicates (each biological replicate is a unique colony from the same Agar plate containing the same gene

module) of the GFP data acquired by the Wallac 1420 and the Excel software. On average 3-5 biological replicates were tested for each module. Standard deviation was calculated for each module behavior by using the mean value ($\bar{x}$) from the 3 to 5 tests, the standard deviation was calculated by using the following formula. P-value was carried out to evaluate the significance of each induction test.

$$s = \sqrt{\frac{\sum(x - \bar{x})^2}{n - 1}}$$

Chapter III – Results

III-1. Construction of the logic function modules

To implement the four logic functions in *E. coli* DH5αZ1, a number of plasmids harbouring novel synthetic genes had to be constructed. We used site-directed mutagenesis (SDM) to modify existing plasmids to create synthetic promoters responding to two input variables (A, IPTG and B, anhydro-tetracycline). Successful SDM was first determined for each synthetic gene min-terms by doing colony PCR with a forward primer specifically targeting only the mutated region and a reverse primer binding to the GFP region or BioBricks suffix region from colonies grown in LB medium after transformation into DH5αZ1. The results of the colony PCR experiments for min-terms 1 to 4 are shown below (Figure 12-15).

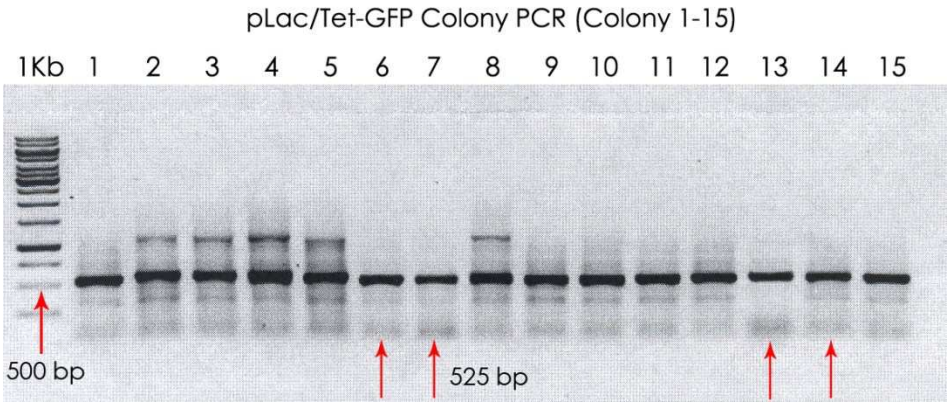

Figure 12. Colony PCR results show a 525 bp band for colonies with successful SDM for min-term1 version 1-pLac/Tet-GFP. The red arrows indicate the colonies that were selected for plasmid sequencing.

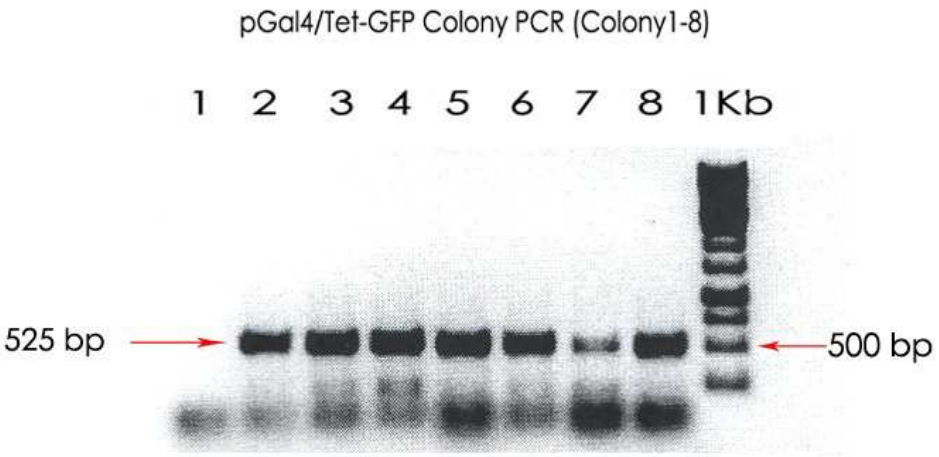

Figure 13. Colony PCR for min-term 2 (A&B'). The results show a 525 bp band for colonies with successful SDM.

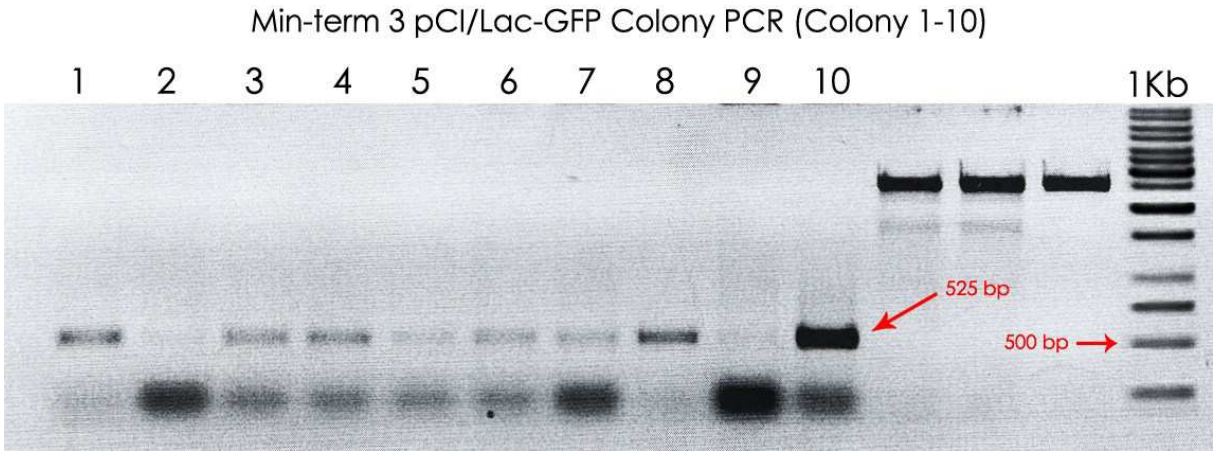

Figure 14. Colony PCR for min-term 3 (A'&B). The band at 525 bp identifies colonies with potential successful SDM.

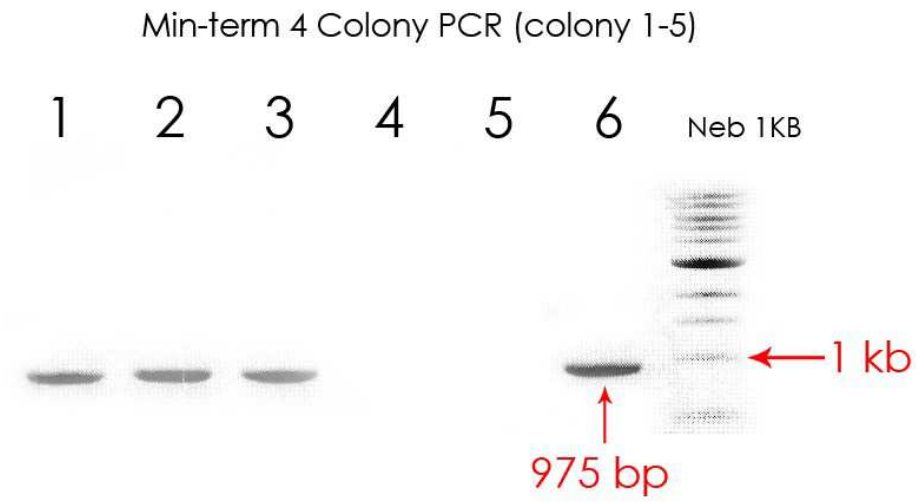

Figure 15. Colony PCR for min-term 4 (A'&B'). The results show the expected unique band at 975 bp. This colony PCR used a different reverse primer, which targets the BioBricks suffix region; the forward primer was still targeting the mutated region in the promoter only. Hence the length of this PCR product is much longer than the previous ones (Figure 12-14).

In addition to the min-terms, two inverter genes had to be constructed to implement the A-B', A'-B and A'-B' logic functions. The same experimental approach described above was used to confirm the construction of the two inverter genes. The first inverter was constructed by the fusion of the lac promoter with the coding sequence of the DNA binding domain of the Gal 4 gene from *S. cerevisiae*. Figure 16 shows the

results of the PCR amplification product of inverter 1 (plac-Gal4) on an 1% Agarose gel. As expected, a 780 bp band is observed.

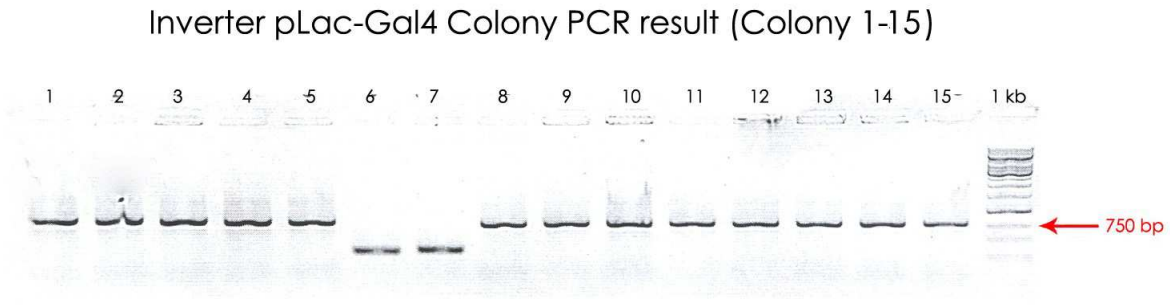

Figure 16. Colony PCR results for the inverter pLac-Gal4 with a terminator sequence. Two primers designed to target the BioBricks prefix and suffix region amplify a 780 bp DNA band corresponding to the complete pLac-Gal4-terminator inverter.

The second inverter was constructed by the fusion of the Tet promoter with the coding sequence of the CI repressor from bacteriophage λ. Figure 17 shows the results of the PCR amplification product of inverter 2 (pTet- λCI) on a 1% agarose gel. As expected, a 913 bp band is observed.

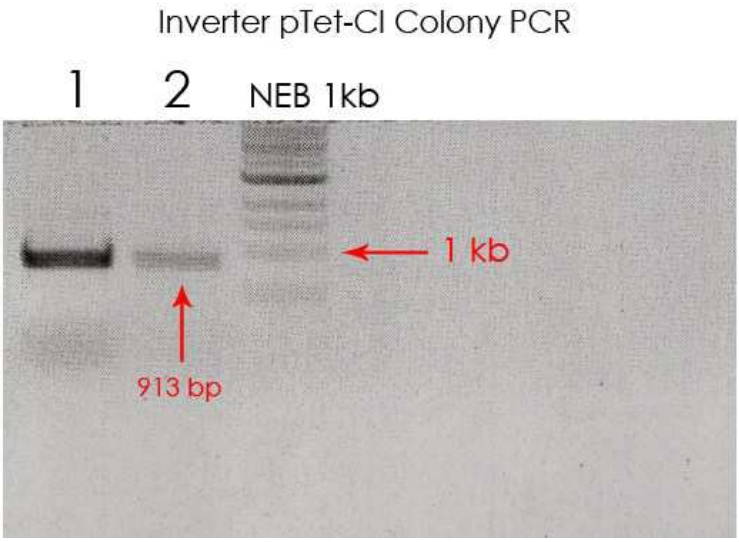

Figure 17. Colony PCR result for the inverter pTet-λCI. Two primers designed to target the BioBricks prefix and suffix region amplify a 913 bp DNA band corresponding to the complete pTet-λC1-terminator inverter.

To implement the A&B', A'&B and A'&B' logic functions, we had to construct the following three plasmids:

pModule2: Fusion of min-term 2 with inverter 1

pModule 3: Fusion of min-term 3 with inverter 2

pModule 4: Fusion of min-term 4 with inverter 1 and 2.

The three plasmids were constructed by the ligation in plasmid pSB1AK3 of the overlapping PCR products of min-term 2, 3 or 4 with the appropriate inverter(s).

Figure 18 shows the results of the PCR products obtained after amplification of the final plasmids with the appropriate primer pairs. As expected, a 1733 bp band is observed for the PCR amplification of module 2 (A-B') corresponding to the fusion of inverter 1 with min-term 2.

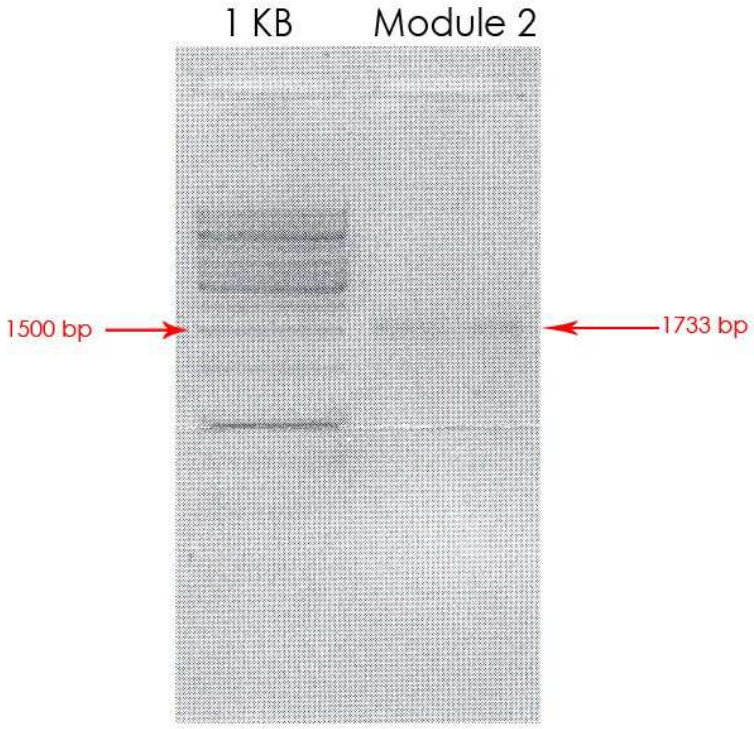

Figure 18. Colony PCR result for pGal4/Tet-GFP and pLac-Gal4 fusion corresponding to the complete Module 2. Two primers targeting the BioBricks prefix and suffix regions were used to amplify the expected 1733 bp region.

Figure 19 shows the results of the PCR amplification of plasmids harbouring the complete module 3. As expected, a 1814 bp DNA band is observed corresponding to

the fusion of min-term 3 with inverter 2.

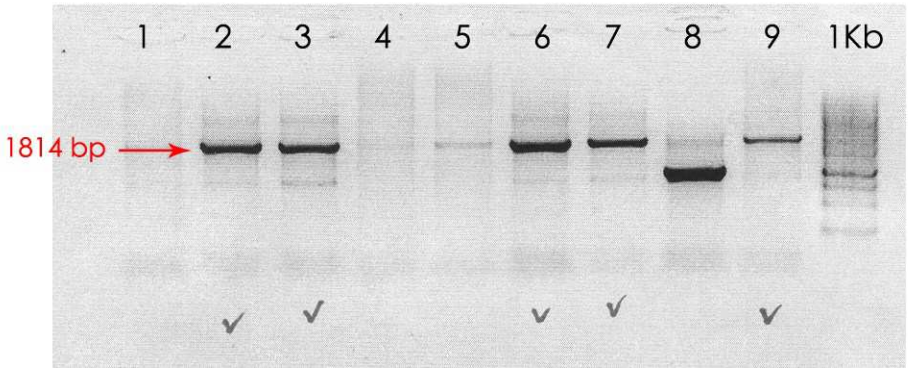

Figure 19. Colony PCR result for pλC1/Lac-GFP and pTet-λC1 fusion corresponding to the complete Module 3. Two primers targeting the BioBricks prefix and suffix regions were used to amplify the expected 1814 bp region.

Finally, a 2529 bp band is observed for the PCR amplification of Module 4 (A'&B') corresponding to the fusion of inverter 1 and 2 with min-term 4 (Figure 20).

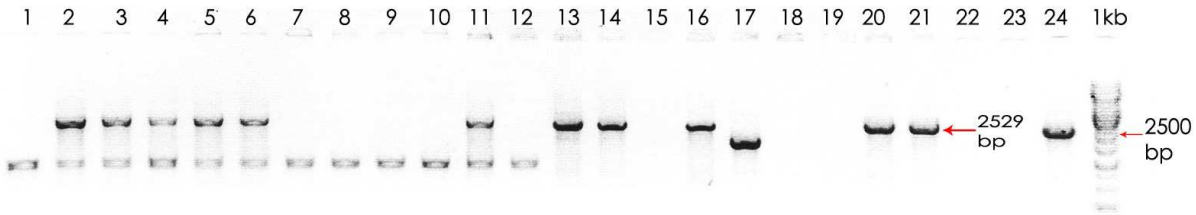

Figure 20. Colony PCR results for the pλCI/Gal4-GFP, pTet-λCI and pLac-Gal4 fusion corresponding to the complete Module 4. Two primers targeting the BioBricks prefix and suffix regions were used to amplify the expected 2529 bp region.

We constructed a total of 10 plasmids; 4 for min-term1, 1 for min-term 2, 1 for min-term 3, 1 for min-term 4, 1 for the pLac-Gal4 inverter and 2 for the pTet-λCI inverter. For each one of them, the entire synthetic construct was sequenced to confirm the modifications that were introduced in the novel synthetic promoters as well as to confirm that no mutations were introduced in the coding sequence of the repressors and of the GFP reporter gene.

Colonies harbouring the proper logic modules were used to test the response of the new synthetic genes to the two input variables, (IPTG and anhydro-tetracycline) *in vivo.* The response was quantified by evaluating GFP fluorescence and compared with the expected patterns for the four different Boolean logic functions implemented in module 1 to 4. For each logic module, 4 biological replicates and 3 technical replicates were performed. The GFP fluorescence values were normalized using the absorbance at 600 nm to control for differences in *E. coli* growth between cultures.

III-2. Evaluation of the behaviour of the logic function modules

Module 1

The four versions of module 1 were tested *in vivo* to see if the pattern of GFP expression corresponds to the expected AND logic gate. The bar charts of the input/output response for all four versions of module 1 are shown in figure 21 to 24. The results show that version 1 (pLac/Tet-GFP) exhibits a 10 to 12-fold increase in

GFP florescence when both inducers are present (with both input variables IPTG and anhydro-tetracycline) compared with any of the other three combinations (no input, single IPTG input or single anhydro-tetracycline input) (Figure 21). Table 1 summarizes the expected theoretical logic response and the actual experimental values obtained with the version 1 of module 1. The version 3 of module 1 also exhibits the expected behaviour (Figure 23). However, the background fluorescence in the un-induced control is higher than the one observed with version 1. In contrast, version 2 and 4 show high background fluorescence and partial induction in presence of a single inducer (Figure 22 and 24).

Module 1 version 1								
Theoretical Response					*In vivo* Response			
	aTc	IPTG	GFP		aTc	IPTG	GFP 1hr	GFP 2 hr
	0	0	0		0	0.24 ng/ml	40833	34107
	1	0	0		0.03 ng/ml	0	41171	35221
	0	1	0		0	0.24 ng/ml	49285	49681
	1	1	1		0.03 ng/ml	0.24 ng/ml	316816	511583

Table 1. Theoretical and *in vivo* response of the version 1 of module 1 to the two inputs: IPTG and anhydro-tetracycline.

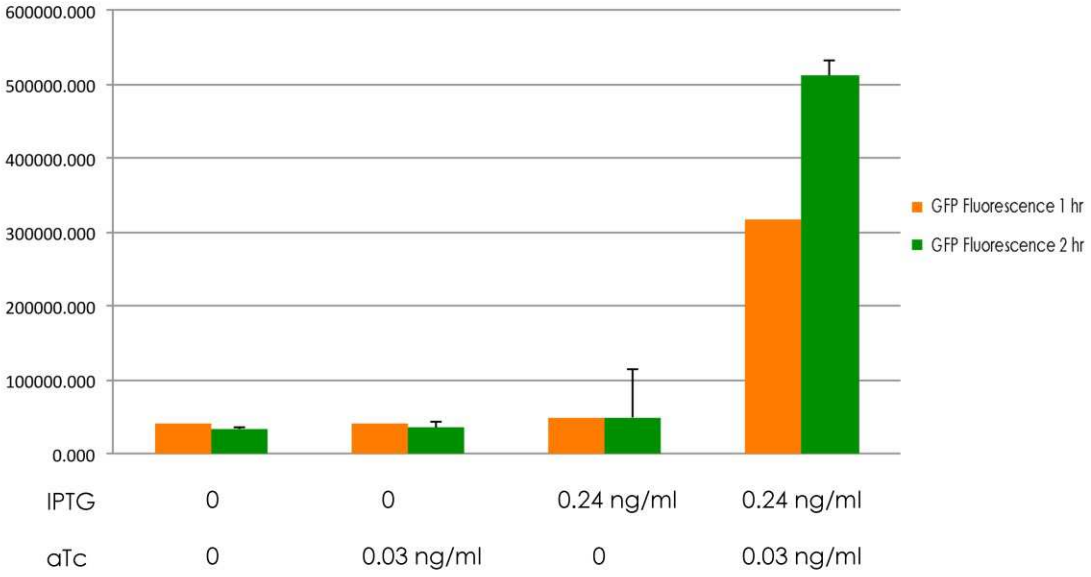

Figure 21. Induction response of module 1-version 1 (pLac-Tet-GFP). GFP output is shown on the Y-axis. GFP fluorescence was recorded one hr (red) or 2 hr (green) after induction

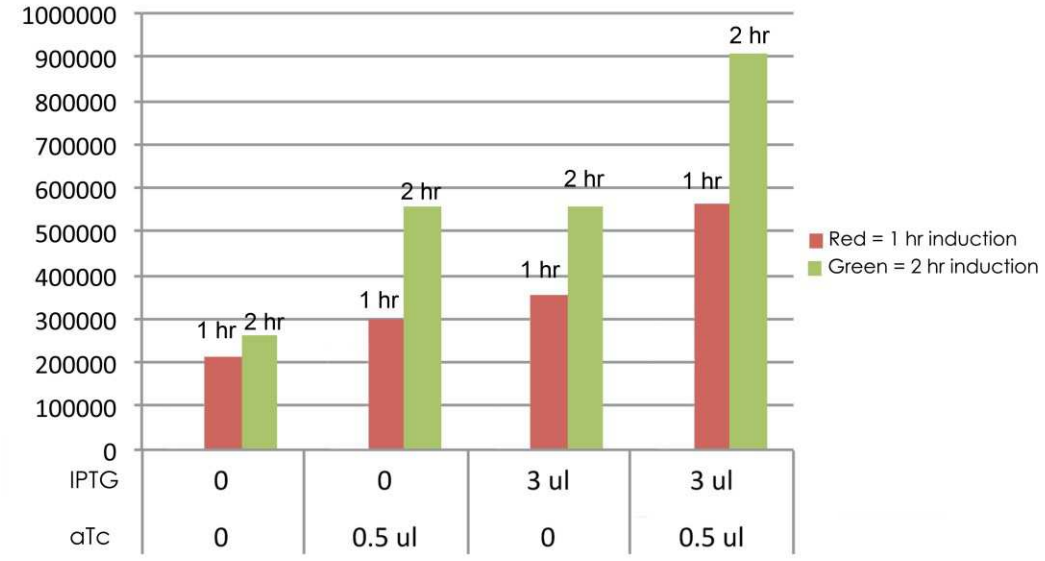

Figure 22. Induction response of module 1-version 2 (pTet-Lac-GFP). GFP output is shown on the Y-axis. GFP fluorescence was recorded one hr (red) or 2 hr (green) after induction

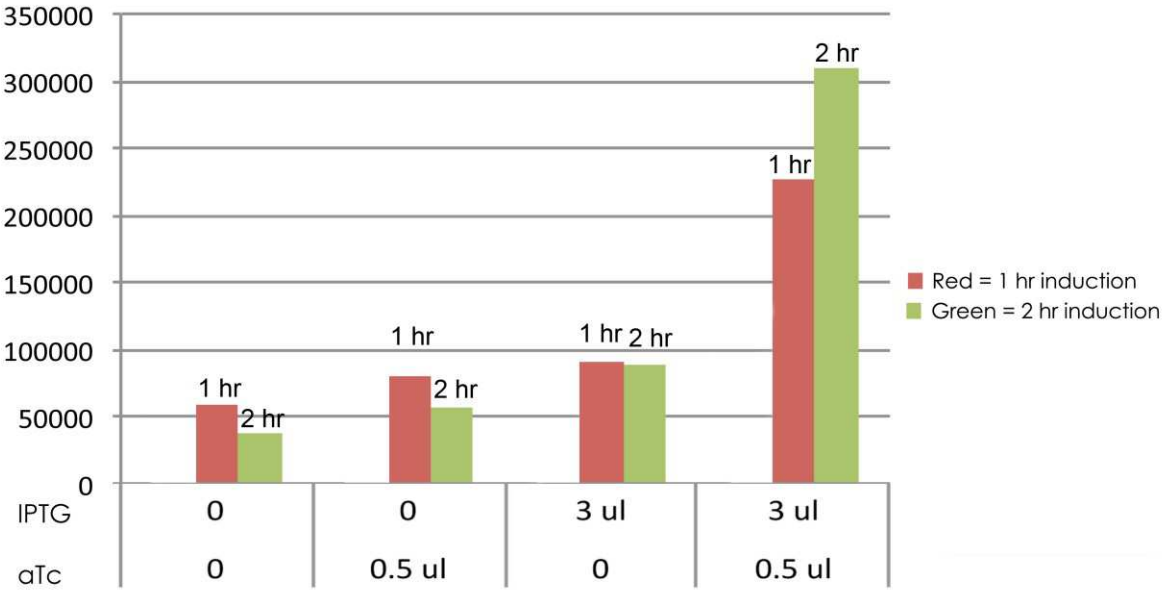

Figure 23. Induction response of module 1-version 3 (pTet-Lac04-GFP). GFP output is shown on the Y-axis. GFP fluorescence was recorded one hr (red) or 2 hr (green) after induction

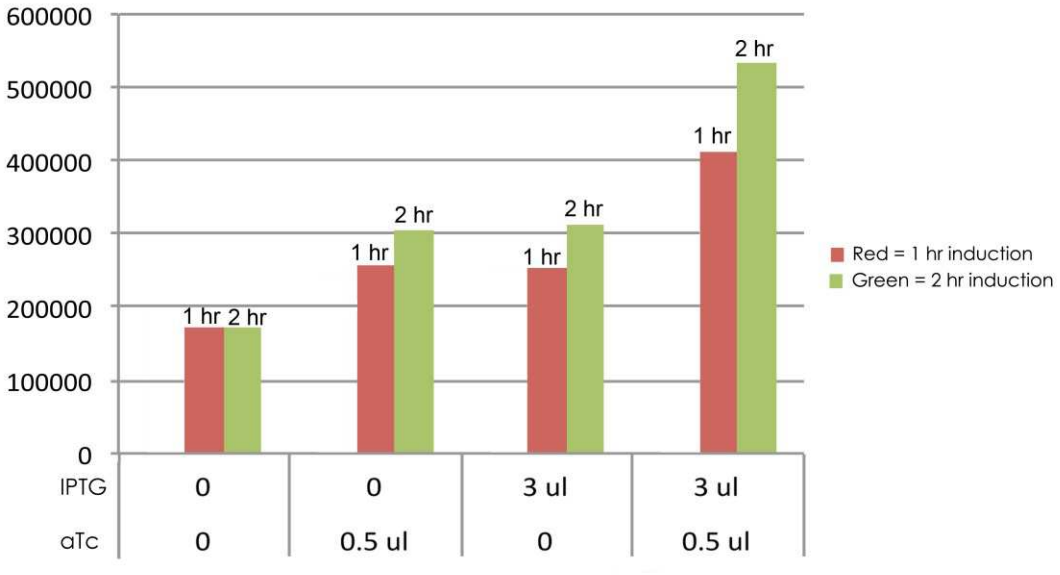

Figure 24. Induction response of module 1-version 4 (pTet-Lac04S-GFP). GFP output is shown on the Y-axis. GFP fluorescence was recorded one hr (red) or 2 hr (green) after induction.

Module 2

Module 2 is made of two genes, pGal4/Tet-GFP and pLac-Gal4 and was created in two steps. First, we created the min-term pGal4/Tet-GFP. The promoter controlling GFP expression is made up of two operators, Tet O1 (between -10 and -35) and one Gal4 binding sequence (right upstream of -35). The second part of module 2 is the inverter gene pLac-Gal4. In this inverter, the sequence coding for the DNA binding domain of Gal4 is placed downstream of the Lac promoter from *E. coli*. The two parts of this module were separately inserted in two individual pSB1AK3 plasmids with the same BioBricks prefix and suffix. Then the two parts were ligated and integrated into one pSB1AK3 plasmid using the BioBricks standard assembly

method [42]. Min-term 2 was placed upstream of the inverter gene and the two parts were separated by a 20bp spacer. The complete module 2 was sent for sequencing and after validation of the sequence it was tested *in vivo* in *E. coli* DH5αZ1 for its response to the two chemical inputs. Table 2 shows the expected theoretical logic response and the actual experimental values obtained with module 2. Figure 25 shows a graphical representation of the results to better appreciate the differences in GFP expression in response to the different treatments. A 2.94 fold difference is observed between the background values and the maximal expected value when only anhydro-tetracycline is present.

Module 2								
Theoretical Response					*In vivo* Response			
	aTc	IPTG	GFP		aTc	IPTG	GFP 1hr	GFP 2 hr
	0	0	0		0	0.24 ng/ml	17192	18531
	1	0	1		0.03 ng/ml	0	28631	54519
	0	1	0		0	0.24 ng/ml	17134	17396
	1	1	0		0.03 ng/ml	0.24 ng/ml	18670	20013

Table 2. Theoretical and *in vivo* responses of module 2 to the addition of inducers.

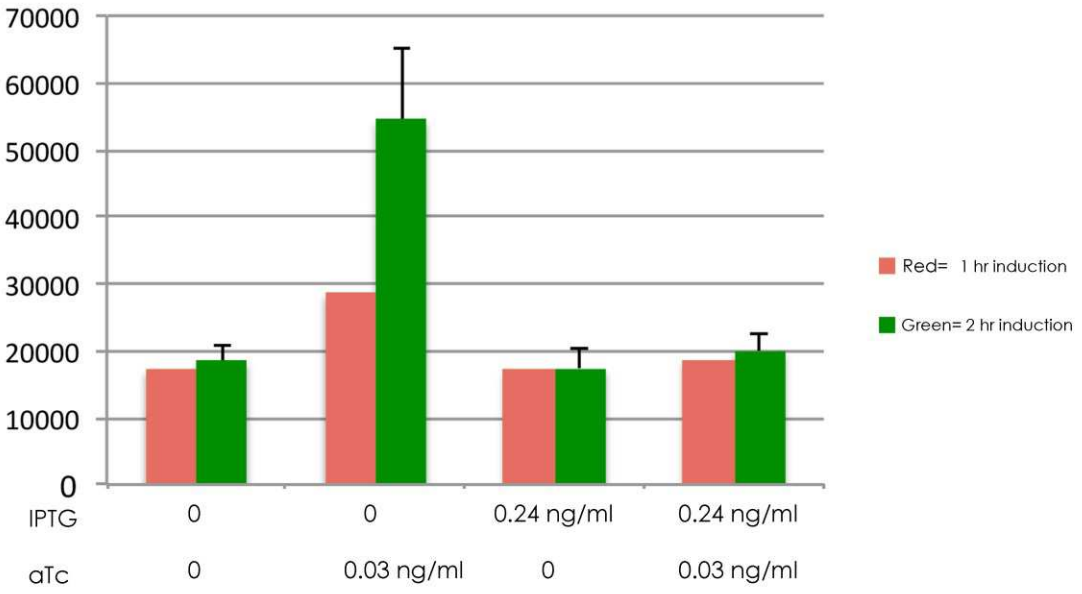

Figure 25. Induction response of module 2. GFP output is shown on the Y-axis. GFP fluorescence was recorded one hr (red) or 2 hr (green) after induction.

Module 3

Following the construction and sequence validation of module 3, the induction test was done to evaluate its performance, The chimeric promoter of module 3, houses the Lac O1 operator and two C1 operators. The theoretical and *in vivo* results are summarized in table 3.

Module 3								
Theoritical Response					*In vivo* Response			
	aTc	IPTG	GFP		aTc	IPTG	GFP 1hr	GFP 2 hr
	0	0	0		0	0.24 ng/ml	128197	145673
	1	0	0		0.03 ng/ml	0	126275	141734
	0	1	1		0	0.24 ng/ml	145648	189713
	1	1	0		0.03 ng/ml	0.24 ng/ml	143258	192317

Table 3. Induction response of module 3. GFP output is shown on the Y-axis. GFP fluorescence was recorded one hr (red) or 2 hr (green) after induction.

Figure 26 shows a graphical representation of the results to better appreciate the differences in GFP expression in response to the different treatments.

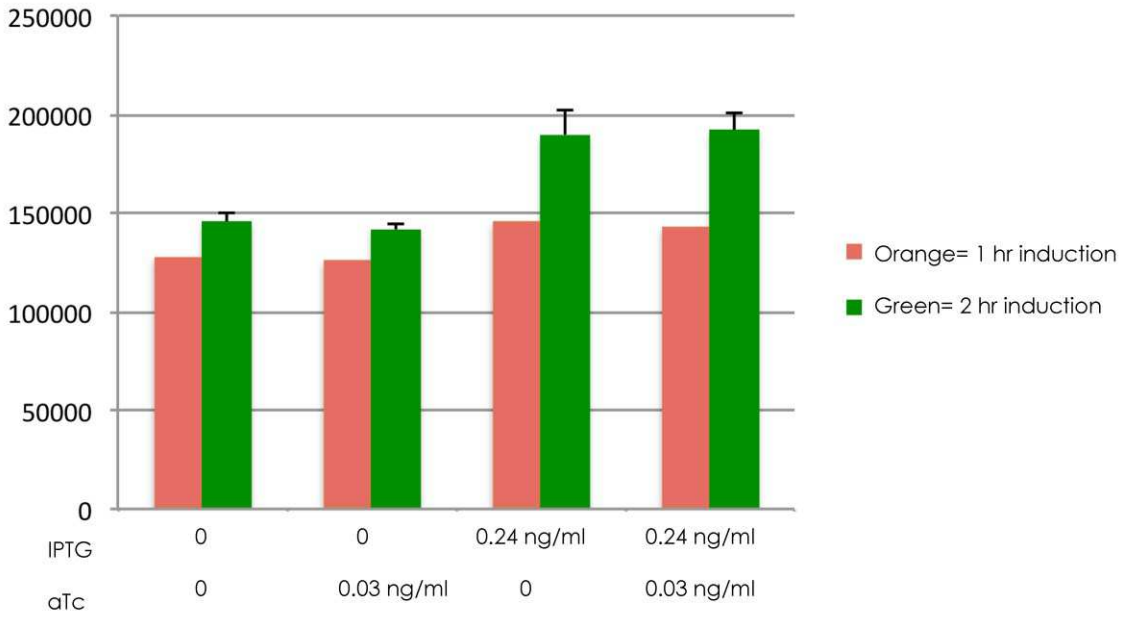

Figure 26. Module 3 *in vivo* induction response test

After 2 hrs of induction, the levels of GFP from the third (IPTG) and fourth treatments (Both IPTG and aTc) showed higher GFP fluorescence than the other two conditions (Table 3, Figure 26). However, the fourth condition should give a GFP fluorescence value similar to treatments 1 and 2. Furthermore, the difference between the GFP values for the fourth treatment is not significantly different from the other treatments. These results might be explained by the high background fluorescence observed when both inducers are absent.

Module 4 is made up of min-term 4 λ pRCl/Gal4-GFP and two inverters, namely pTet-λCl and pLac-Gal4. All three genes were constructed separately and ligated in plasmid pSB1AK3 in the following order pCl/Gal4-GFP, pTet-λCl and pLac-Gal4. A 20bp spacer was introduced between the three genes. Table 4 shows the expected theoretical logic response and the actual experimental values obtained with module 4. Figure 27 shows a graphical representation of the results to better appreciate the differences in GFP expression in response to the different treatments.

Module 4								
Theoretical Response					*In vivo* Response			
	aTc	IPTG	GFP		aTc	IPTG	GFP 1hr	GFP 2 hr
	0	0	1		0	0.24 ng/ml	109764	112913
	1	0	0		0.03 ng/ml	0	104787	88412
	0	1	0		0	0.24 ng/ml	112465	110048
	1	1	0		0.03 ng/ml	0.24 ng/ml	108147	92316

Table 4. Theoretical and *in vivo* response of module 4 to the addition of inducers

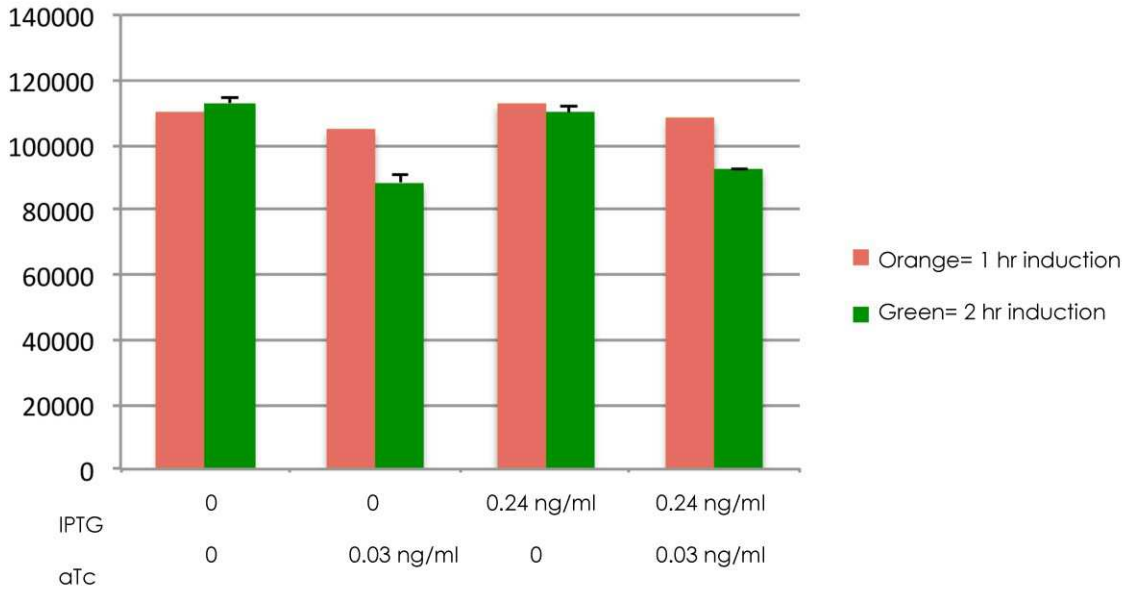

Figure 27. Induction response of module 4. GFP output is shown on the Y-axis. GFP fluorescence was recorded one hr (red) or 2 hr (green) after induction.

Initially, min-term 4 λ pRC1/Gal4-GFP was constructed using λ pR-GFP (BBa_S03335) from the BioBricks Part Registry as starting material. SDM was carried out to insert the Gal4 Binding sequence between the SD sequence and the transcription start site. Colony PCR was used to screen for positive colonies harbouring the mutant plasmid. Then this min-term was ligated to pTet-λC1 in pSB1AK3. This partial construct was tested *in vivo* and the results showed strong repression mediated by the λ C1 repressor even in the absence of induction by anhydro-tetracycline (Table 5). The results suggest that the small amount of λ C1 produced by the leaky pTet promoter is sufficient to repress GFP production.

Min-term 4 without inverter pTet- λCI					Min-term 4 with inverter pTet- λCI		
	Colony1	Colony2	Colony3		Colony 1	Colony 2	Colony 3
2 hr	23866	23536	23776		9813	8216	8528

Table 5. GFP production from min-term 4 in the absence of inverter (left 3 columns) and in the presence of the inverter (right 3 columns)

To decrease the amount of C1 produced from the inverter pTet-λCl gene, a weaker version of the promoter was created by reducing the strength of the RBS to 11.9% of its original strength. [43] This new version of pTet-λCl was constructed via SDM and integrated into the complete module 4. Upon finishing constructing module 4 the module was tested for its logic response.

III-3. Statistic tests for module 1 and 2

Student t-test was calculated to evaluate the significance of the mean values for the induction tests. The t-tests were done for each of the theoretical false value state (0) against the true value state (1). The results and sample sizes are summarized in the following table. Module 3 and 4 t-tests were not done here because the differences between the false states and the true states were not significant.

Module 1	t-test			Sample size
Between	no induction	aTc only	IPTG only	
Double induction	< 0.0001	< 0.0001	< 0.0001	3
Module 2	t-test			
Between	no induction	IPTG only	Double Induction	
aTc only	0.0236	0.022	0.0345	3

Table 6. Student t-test for module 1 and 2

Chapter IV – Discussion

The purpose of the present project is to construct an *E. coli* cell with a configurable information processing functionality. Ultimately, the cell will have the ability to perform any of 16 different 2-input combinatorial logic functions. The design is based on the construction of genes responding to two input chemicals, IPTG and anhydro-tetracycline, and producing one output, the green fluorescent protein. In our design, the output protein is expressed only if a particular logic function of the two input chemicals is true. The implementation of this information processing functionality requires two steps.

1) The construction of 6 synthetic genes organized in 4 different logic modules. Two of these genes (inverters) are involved in more than one module while the four others (min-terms) are exclusive to their respective module (Figure 2).
2) Their introduction in the chromosome of *E. coli*.

Once completed, the synthetic cell can be configured to perform any one of the 16 different Boolean logic functions with two input variables. The configuration step will require the excision from the *E. coli* genome of one or more of the 4 logic modules to create all possible combinations.

My contribution to this project was to construct the 6 genes and test the four modules in *E. coli*.

Module 1 [A&B]

We successfully constructed the "A&B" module, which exhibited a 10-fold difference in GFP expression in presence of the two inducers, compared with the other three possible combinations (figure 21). This behaviour is in line with an "AND" logic gate behavior. This module required the construction of a single gene [min-term 1] harboring a modified Tet promoter. Although synthetic "AND" gates have been successfully created in the past, their design were based on the use of more complex promoters. For example, a previously described "AND" promoter responding to IPTG and arabinose contained the 3 operators from the naturally occurring Lac operon and 2 operators from the arabinose operon in addition to the cAMP-CAP binding site of the Lac operon [16]. For the first logic module of this project, our goal was to create a minimal "AND" gate promoter using only two operator sequences. Therefore we constructed different variants where only one Lac operator and one Tet operator would regulate gene expression. For such a promoter, leakiness is expected since we are reducing the amount of regulatory sequences and the results indeed suggested that GFP was being produced even in the absence of inducers. However, the 10-fold difference in GFP fluorescence observed when both inducers are present as compared with the other possible combinations is high enough to allow its use as a synthetic "AND" gate. Furthermore, we can expect to observe greater differences between induced and non-induced levels when min-term 1 will be introduced as a single copy in the chromosome.

To try to improve the behaviour of min-term 1 and reduce background expression in absence of inducers, we constructed different variants where the operator sequence and/or its localization in the promoter was changed. In the first attempt (variant 2), we kept the TetR1 operator upstream of the -35 box and replaced the TetR2 which is normally located between the -35 and -10 boxes with the Lac O1 operator (Figure 5). We were expecting to observe a stronger repression in absence of IPTG since this localization was previously shown to provide a tighter repression of the chimeric Lac-ara promoter [16]. Surprisingly, this configuration resulted in higher background expression values with less than two-fold differences between the induced (A&B), partially induced (A or B) and un-induced conditions (Figure 22). The differences observed between our construct and the chimeric Lac-Ara promoter might be explained by the fact that we used the Tet promoter instead of the Lac promoter for the introduction of the novel operator sequence and that the sequence context into which the operator is located is important for its impact on repression. In the third and fourth versions of module 1, we introduced different Lac operator sequences between the -35 and -10 boxes to see if the background levels in absence of IPTG would be reduced. We selected the Lac operator sequences lacO4 and Lac04S to which LacI was found to bind with higher affinity [41]. When comparing variant 1 with variant 3 and 4, we can see that although the Lac operator used in version 1 is relatively weak compared with the two operators used in version 3 and 4, the overall performance of the version 1 was still better. Again, the difference can be explained by the position of the operator, which is different between the

constructs. It would be interesting to test another variant having LacO4 or LacO4s located upstream of the -35 box as in variant 1.

Module 2

We successfully constructed the "non A&B" module, which exhibited a 3-fold difference in GFP expression in the absence of IPTG and in presence of anhydro-tetracycline compared with the other three possible combinations (figure 25). Module 2 is composed of two genes, the min term pGal4/Tet-GFP and the inverter pLac-Gal4 to negate the first variable, IPTG. The pGal4/Tet-GFP min-term was constructed using the same design as the one used for the first variant of min-term1. The only difference being that the lacO1 operator present in min-term 1 was replaced with the yeast UASg sequence. This sequence is recognized by the Gal4 protein which upon binding acts as repressor of transcription in bacteria. It has been shown previously that a single UASg sequence introduced in the Lac promoter at the Lac O1 operator position could repress the Lac promoter by 30-fold in presence of Gal4 [44]. In the absence of IPTG, the min-term is repressed by the Tet repressor only while Gal4 is expressed and acts as a repressor only when IPTG is added to the culture. In our construct, a three-fold difference is observed between the fully induced state (anhydro-tetracycline only) and the Gal4 or TetR repressed state. The Tet promoter is generally regarded as a stronger promoter compared to the Lac promoter; hence the rate of complex formation between the RNA polymerase and the promoter is faster (faster K_{on}). In a previous study, it has been shown that there is an inverse correlation between the K_{on} and promoter repression

[41,46]. The position of the Gal4 binding sequence in module 2 is upstream and slightly overlapping with the -35 box whereas in the study that showed the 30-fold repression mediated by Gal4, the position of the Gal4 binding sequence was placed at the original position of Lac O1 in the weaker lac promoter. In future attempts, different promoter backbones with the UASg at different positions can be designed and tested to maximize the repression mediated by Gal4 or TetR.

Surprisingly, the maximal activity observed with min-term 2 is about 10 times lower than the one reached with min-term 1 in the fully induced state. The lower activity of the min-term 2 promoter must be caused by the introduction of the UASg sequence since it is the only difference between min-term 1 and min-term 2. Even though it is generally assumed that it is the "quality" of the -35 and -10 boxes that determines promoter strength in prokaryotes, several studies have shown that the surrounding sequences have an impact on expression [46]. To improve our results, we will have to evaluate the impact of moving the UASg sequence at other positions in the promoter on GFP expression.

Module 3 and 4

Module 3 and 4 did not exhibit the expected expression behavior. The min-terms of both modules were constructed using as backbone the lambda Pr promoter. In bacteriophage lambda, this promoter drives the expression of the Cro gene and is repressed by the C1 repressor. The selection of this promoter as backbone for both min-terms seemed a rationale choice since we also decided to use C1 as repressor

for both modules. Module 3 is composed of two genes, the min-term pC1/Lac-GFP and the inverter pTet-C1 to negate the second variable, anhydro-tetracycline. High expression should only be observed when IPTG is present and anhydro-tetracycline is absent. Our construct did not exhibit this behavior and showed the same high GFP fluorescence with all the different combinations of inducers. This result is surprising since our chimeric promoter retains the natural C1 binding sites, Or1 and Or2 and should at least be strongly repressed when C1 is expressed. One possible explanation for this unusual behavior is that the introduction of the lac01 operator in the chimeric promoter affects negatively the binding of both the C1 and LacI repressors [45].

A similar unexpected behavior was obtained with module 4. Module 4 is composed of three genes, the min-term pC1/Gal4-GFP and the inverters pTet-C1 and pLac-Gal4 to negate both variables (IPTG and anhydro-tetracycline). The configuration of the operators in min-term 4 is similar to the one used in min-term 3. The synthetic promoter of min-term 4 retains the C1 binding sites Or1 and Or2 at their original position with the additional UASg sequence embedded upstream of the -10 box and downstream of the Shine-Dalgarno sequence. Under natural conditions, this promoter is not repressed since *E. coli* does not produce the Gal4 or C1 protein. However, the introduction of the two inverters should allow the production of the repressors when the appropriate inducer is present. We know that the C1 repressor is produced from the inverter gene as shown by the results of Table 5. We also know that the Gal4 protein is produced from the inverter gene since it was able to repress the expression of module 2. It seems that the combination of the lambda operators

Or1 and Or2 in combination with another operator located downstream of the -10 box interferes with repression.

In the future, we will have to reconstruct module 3 and 4 based on a completely different design. First, our results demonstrate that we have to choose carefully the background promoter. The choice should allow the production of min-terms having similar maximum expression levels and strong repression. Unfortunately, we cannot predict with accuracy the effect of introducing new operator sequences in existing promoters. To by-pass this limitation, we will have to test systematically different promoters, operator sequences and operator sequence configurations to maximize their behavior according to the logic functions that we want to implement.

References

1. Feynman RP. In Minaturization D.H. Gilbert Ed. Reinhold Publishing Corporation

New York, 1961 pp 282-296

2. Szybalski W, *In Vivo* and *in Vitro* Initiation of Transcription, Page 405. In: A. Kohn and A. Shatkay (Eds.), Control of Gene Expression, pp. 23-24, and Discussion pp. 404-405 (Szybalski's concept of Synthetic Biology), 411-412, 415 - 417. New York: Plenum Press, 1974

3. Conference Proceedings, Synthetic Biology 3.0 ETH Zurich, Switzerland, June 24-26 2007

4. Gardner TS, Cantor CR, Collins JJ. Construction of a genetic toggle switch in *Escherichia coli.* Nature. 2000 Jan 20; 403 (6767): 339-42.

5. Kramer, B.P. et al. An engineered epigenetic transgene switch in mammalian cells. Nat. Biotechnol. (2004) vol 22, 867–870

6. Isaacs, F.J., Hasty, J., Cantor, C.r. & Collins, J.J. Prediction and measurement of an autoregulatory genetic module. Proc. Natl. Acad. Sci. 2003 USA 100, 7714–7719.

7. Ham, T.S., Lee, S.K., Keasling, J.D. & Arkin, A.P. A tightly regulated inducible expression system utilizing the fim inversion recombination switch. Biotechnol. Bioeng. 94, 1–4 (2006).

8. Ham, T.S., Lee, S.K., Keasling, J.D. & Arkin, A.P. Design and construction of a double inversion recombination switch for heritable sequential genetic memory. PLoS ONE 3, e2815 (2008).

9. Deans, T.L., Cantor, C.r. & Collins, J.J. A tunable genetic switch based on RNAi and repressor proteins for regulating gene expression in mammalian cells. Cell 130, 363–372 (2007).

10. Ajo-Franklin, C.M. et al. rational design of memory in eukaryotic cells. Genes Dev. 21, 2271–2276 (2007).

11. Friedland, A.E. et al. Synthetic gene networks that count. Science 324, 1199–1202 (2009).

12. Rinaudo, K. et al. A universal RNAi-based logic evaluator that operates in mammalian cells. Nat. Biotechnol. 25, 795–801 (2007).

13. Win, M.N. & Smolke, C.D. Higher-order cellular information processing with synthetic RNA devices. Science 322, 456–460 (2008).

14. Anderson JC, Voigt CA, Arkin AP. Environmental signal integration by a modular AND gate. Mol Syst Biol. 2007;3:133. Epub 2007 Aug 14.

15. T. F. Knight, Jr., G. J. Sussman. Cellular Gate Technology Proc. UMC98, First International Conference on Unconventional Models of Computation, Auckland, NZ, January 1998.

16. Lutz R, Bujard H. Independent and tight regulation of transcriptional units in *Escherichia coli* via the LacR/O, the TetR/O and AraC/I1-I2 regulatory elements. Nucleic Acids Research, 1997, Vol. 25, No. 6 1203–1210

17. Rinaudo K, Bleris L, Maddamsetti R, Subramanian S, Weiss R, Benenson Y. A universal RNAi-based logic evaluator that operates in mammalian cells. Nature Biotechnology. May 2007; doi:10.1038/nbt1307

18. Bayer, T.S. & Smolke, C.D. Programmable ligand-controlled riboregulators of eukaryotic gene expression. Nat. Biotechnol. 23, 337–343 (2005).

19. Kobayashi, H. et al. Programmable cells: interfacing natural and engineered gene networks. Proc. Natl. Acad. Sci. USA 101, 8414–8419 (2004).

20. Win, M.N. & Smolke, C.D. A modular and extensible RNA-based gene-regulatory platform for engineering cellular function. Proc. Natl. Acad. Sci. USA 104, 14283–14288 (2007).

21. Anderson, J.C., Clarke, E.J., Arkin, A.P. & voigt, C.A. Environmentally controlled invasion of cancer cells by engineered bacteria. J. Mol. Biol. 355, 619–627 (2006).

22. Kramer BP, Fischer M, Fussenegger M. Semi-synthetic mammalian gene regulatory networks. Metabolic Engineering 7 241–250 (2005)

23. Basu S, Mehreja R, Thiberge S, Chen MT, and Weiss R, Spatiotemporal control of gene expression with pulse-generating networks. PNAS April 27, 2004 vol. 101 no. 17 6355–6360

24. Wang BJ, Kitney RI, Joly N, Buck M. Engineering modular and orthogonal genetic logic gates for robust digital-like synthetic biology. Nature Communications Article number: 508 doi:10.1038/ncomms1516, 2011

25. Hunziker A, Tuboly C, Horváth P, Krishna S, Semsey S. Genetic flexibility of regulatory networks. PNAS July 20, 2010 vol. 107 no. 29 12998–13003

26. Beckwith, J. in Escherichia coli and Salmonella typhimurium: Cellular and Molecular Biology (ed. Neidhardt, F. C.) 1444–1452 (American Society for Microbiology, Washington DC, 1987)

27. Klig, L. S., Carey, J. & Yanofsky, C. trp repressor interactions with the trp, aroH and trpR operators. Comparison of repressor binding *in vitro* and repression *in vivo*. J. Mol. Biol. 202, 769–777 (1988)

28. Somerville, R. The Trp repressor, a ligand-activated regulatory protein. Prog. Nucleic Acid Res. Mol. Biol. 42, 1–38 (1992)

29. Oehler S, Amouyal M, Kolkhof P, Wilcken-Bergmann B, Muller-Hill B Quality and position of the three lac operators of *E. coli* define efficiency of repression. The EMBO Journal vol.13 no.14 pp.3348-3355, 1994

30. Guet CC, Elowitz MB, Hsing W, Leibler S. Combinatorial Synthesis of Genetic Network. Science 24 May 2002: Vol. 296 no. 5572 pp. 1466-1470 DOI: 10.1126/science.1067407

31. Ashkenasy G, Ghadiri MR. Boolean logic functions of a synthetic peptide network. J Am Chem Soc. 2004 Sep 15;126(36):11140-1.

32. Benenson Y, Gil B, Ben-Dor U, Adar R, Shapiro E. An autonomous molecular computer for logical control of gene expression. Nature. 2004 May 27;429 (6990):423-9. Epub 2004 Apr 28.

33. Kramer BP, Fischer C, Fussenegger M. BioLogic Gates Enable Logical Transcription Control in Mammalian Cells. Wiley InterScience 2004.

34. Mayo A.E., Setty Y.,Shavit S., Zaslaver A., Alon U. Plasticity of the cis-Regulatory Input Function of a Gene. PLoS Biology | April 2006 | Volume 4 | Issue 4 | e45

35. Istrail S., Davidson E.H. Logic functions of the genomic cis-regulatory code. PNAS. Apr 5, 2005 Vol 102 no.14. 4954–4959

36. Cox III R.S., Surette M.G., Elowitz M. B. Programming gene expression with combinatorial promoters. Molecular Systems Biology 3 2007; Article number 145; doi:10.1038/msb4100187

37. Schilstra M.J. , Nehaniv C.L., Bio-Logic: Gene Expression and the Laws of Combinatorial Logic. Artificial Life. Winter 2008, Vol. 14, No. 1, Pages 121-133 doi:10.1162/artl.2008.14.1.121.

38. Buchler N.E., Gerland U, HwaT. On Schemes of Combinatorial Transcription Logic. PNAS April 29, 2003, vol. 100. no. 9. 5136–5141.

39. Ceroni F, Furini S, Giordano E, Cavalcanti S. Rational design of modular circuits for gene transcription: A test of the bottom-up approach Journal of Biological Engineering 2010, 4:14

40. Smolke, C.D., Building outside of the box: iGEM and the BioBricks Foundation. Nature Biotechnology 27, 1099 - 1102 (2009) doi:10.1038/nbt1209-1099

41. Lanzer M Bujard H. Promoters largely determine the efficiency of repressor action. Biochemistry Proc. Natl. Acad. Sci. USA Vol. 85, pp. 8973-8977, December 1988

42. Biobricks Standard Assembly Protocol: http://partsregistry.org/Assembly:Standard_assembly

43. Biobricks RBS library, http://partsregistry.org/Part:BBa_J61101

44. Paulmier N. Yaniv M. Wilcken-Bergman B. V. Muller-Hill B. gal4 transcrition activator protein of yeast can function as a repressor in *Escherichia Coli*. 1987 EMBO Journal vol 6 no.11 pp 3539-3542

45. Brenowitz M, F.Senear D. and K.Ackers G. Flanking DNA-sequences contribute to the specific binding of cI-represor and OR1 Nucleic Acids Research 1989 Volume 17 Number 10

46. Kedracka-Krok S. and Wasylewski Z. Kinetics and Equilibrium Studies of Tet Repressor–Operator Interaction JOURNAL OF PROTEIN CHEMISTRY 1999 Volume 18, Number 1, 117-125